

Lecture 21 - scribbles

Tuesday, November 20, 2018 14:34

today: continue MCMC
Gibbs sampling & M.H.

continue Markov chain review:

def: **irreducible Markov chain** \Leftrightarrow there exists a positive prob. "path"
from every $i \rightarrow j$
 $\forall (i, j), \exists$ an integer m_{ij} s.t. $(A^{m_{ij}})_{ij} > 0$

(by Perron-Frobenius thm.) \Rightarrow irreducible M.C. has a unique stationary dist.

* in order to converge to it, we need aperiodicity as well

irreducible and aperiodic M.C. $\Leftrightarrow \exists$ an integer m s.t. $A^m > 0$
(i.e. $(A^m)_{ij} > 0 \forall i, j$)
aka. regular M.C. (finite state)
or ergodic M.C.

(*) [note: a sufficient condition for an irreduc. M.C. to be aperiodic is $\exists i$ s.t. $A_{ii} > 0$]

$$A = \begin{pmatrix} 0 & 1/2 & 1/2 \\ 1/2 & 0 & 1/2 \\ 1/2 & 1/2 & 0 \end{pmatrix} = \frac{1}{2}(\mathbb{1}\mathbb{1}^T - I)$$

12 1 constant ... +1

$$A^2 = \frac{1}{4} \begin{pmatrix} 1/2 & 0 & 1/2 \\ 1/2 & 1/2 & 0 \end{pmatrix} \dots$$

$$A^2 = \frac{1}{4} (\underbrace{AA^T}_{K} - 2AA^T + I)$$

$$= \frac{1}{4} ((K-2)AA^T + I)$$

thm: if a finite M.C. is ergodic (regular)

then \exists a unique stationary dist. π

and for any starting dist. π_0 , $\lim_{t \rightarrow \infty} A^t \pi_0 = \pi$

the speed of convergence is related to the mixing time τ of the chain

$$\tau \triangleq \frac{1}{1 - \lambda_2(A)}$$

λ_2 2nd biggest e-value of A

$$\| A^t \pi_0 - \pi \|_1 \leq C \exp(-t/\tau)$$

* intuition (from linear algebra) [informal argument]

simpler case, suppose A is diagonalizable with orthogonal matrix U
is symmetric

$$A = U \Sigma U^T \text{ with } \Sigma = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_K \end{pmatrix}$$

U \rightarrow basis of e-vectors

(by Perron-Frobenius thm.)

$U = (u_1 \dots u_k)$ take $u_1 = \pi / \|\pi\|_2$
 $[A\pi = \pi]$

can show that $|\lambda_1| = 1 > |\lambda_2| \geq \dots \geq |\lambda_k|$

let α_0 be coordinate in basis U st.

$$U^T U = I = U U^T$$

$$\pi_0 = U \alpha_0 \quad (\text{ie. } \alpha_0 = U^T \pi_0 \quad (\alpha_0)_1 = \frac{\pi^T \pi_0}{\|\pi\|_2})$$

$$A^t \pi_0 = (U \Sigma U^T)^t (U \alpha_0) = U \Sigma^t U^T U \alpha_0$$

$$= U \Sigma^t \alpha_0$$

$$\Sigma^t = \begin{pmatrix} 1 & & & \\ & \lambda_2^t & & \\ & & \dots & \\ & & & \lambda_k^t \end{pmatrix}$$

$$A^t \pi_0 = \left[(\alpha_0)_1 \cdot \underbrace{\frac{u_1}{\|\pi\|_2}}_{\rightarrow 0} + (\alpha_0)_2 \lambda_2^t u_2 + \dots + (\alpha_0)_k \lambda_k^t u_k \right]$$

because $1^T A^t \pi_0 = 1 \quad \forall t \Rightarrow (\alpha_0)_1 = 1$ [not fully correct?]

$$\|A^t \pi_0 - \pi\|_2 \leq C |\lambda_2|^t$$

$$|\lambda_2| = 1 - \epsilon_1 \quad \epsilon_1 \triangleq 1 - |\lambda_2|$$

$$|\lambda_2| \leq \exp(-\epsilon_1) \quad 1 - x \leq \exp(-x) \quad \forall x$$

$$|\lambda_2|^t \leq \exp(-t \epsilon_1)$$

$$\Rightarrow \tau = \frac{1}{1 - |\lambda_2|}$$

* mixing time is often exponentially big

⊗ How do we design A st. $A^t \pi_0 \rightarrow \pi_0$?

one "easy way"

reversible M.C. \Leftrightarrow iff \exists dist. π s.t. $A_{ij}\pi_j = A_{ji}\pi_i \quad \forall (i,j)$

"detailed balance equation"

it means that $P\{X_t=i, X_{t-1}=j\} = P\{X_t=j, X_{t-1}=i\}$
(when $P\{X_{t-1}=i\} = \pi_i$)

sufficient condition (but not necessary)

for $A\pi = \pi$

proof: $(A\pi)_i = \sum_j A_{ij}\pi_j \stackrel{\text{detailed balance}}{=} \sum_j A_{ji}\pi_i = \pi_i \left(\sum_j A_{ji} \right) //$

15h25

Metropolis-Hastings alg.

goal \rightarrow construct a M.C. with stationary dist. $p(x)$ [our target]

(assume $p(x) > 0 \quad \forall x$)

uses some proposal $q(x'|x)$

[i.e. if in state x , we sample $x'|x \sim q(x'|x)$]

depends on $\frac{p(x')}{p(x)}$, no need to normalize

ϕ
run

[ie. if in state x , we sample $x'|x \sim q(x'|x)$]

accept new state x' with prob.

if reject \rightarrow stay in same state x

[this still a new sample]

vs. rejection sampling where only "accepted states" are sample

$$a(x'|x) \triangleq \min \left\{ 1, \frac{q(x|x') p(x')}{q(x'|x) p(x)} \right\}$$

acceptance ratio to satisfy detailed balance

MH alg.:

start at $x^{(0)}$

for $t=1, \dots$

• propose $x^{(t)} \sim q(x'|x^{(t-1)})$

• flip a biased coin with prob $a(x^{(t)}|x^{(t-1)})$ to be 1

• if accept (coin = 1):

let $x^{(t)} = x^{(t)}$

o.w.

$x^{(t)} = x^{(t-1)}$

end for

note: for symmetric $q(x'|x)$, always accept if $p(x') \geq p(x)$

[Metropolis alg.]

\rightarrow like a noisy hill-climbing alg.

\swarrow target

[Metropolis alg.]

[verify as exercise that it satisfies detailed balance]

with $\pi = p$ \swarrow target

* for convergence: if M.H. chain is ergodic, then we converge to correct unique stationary dist. P

sufficient conditions $\left\{ \begin{array}{l} \text{for irreducibility } q(x'|x) > 0 \quad \forall x' \neq x \in X \\ \text{for aperiodicity } q(x|x) > 0 \text{ for some } x \in X \end{array} \right.$

(*) aside: it is still ok to change proposal with time
(inhomogeneous M.C.) $q_t(x'|x)$

as long as choice of q_t does not depend on $x^{(t-1)}$

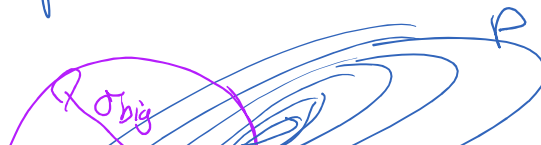
then convergence theory above will go through

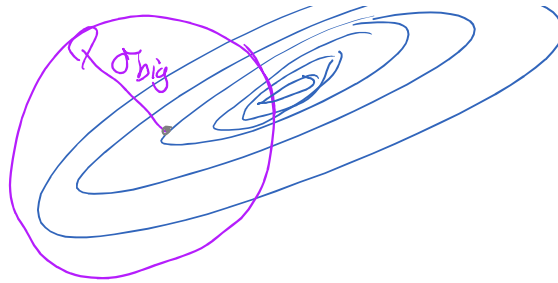
[i.e. detailed balance, etc. will give right stationary dist.]

slow mixing example:

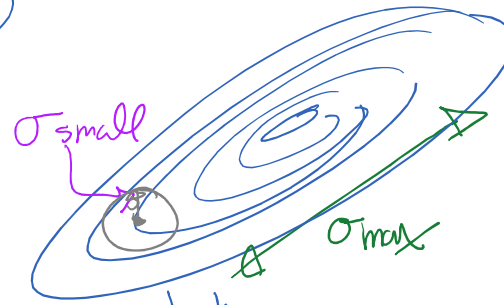
suppose p is a multivariate normal

$$\xi \quad q(x'|x) = N(x'|x, \sigma^2 I)$$





* high proba. of rejection



here the ^{best.} mixing time is related to ratio $\frac{\sigma_{max}}{\sigma_{min}}$

good book: Casella & Berger

"Monte Carlo Statistical methods"

Gibbs sampling alg.

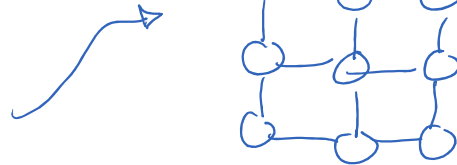
↳ M.H. with a clever choice of proposal $q_t(x'|x)$

examples of applications:

↳ UGM: $\tilde{p}(x) = \prod_c \psi_c(x_c)$

↳ difficult conditional in DGM

$\tilde{p}(x) = p(x, \bar{x}_E) \delta(x_E, \bar{x}_E) \propto p(x | \bar{x}_E)$



UGM:

cyclic Gibbs sampling alg.: nodes $i = 1, \dots, n$

start at some $x^{(0)}$

run $\downarrow - 1$

start at some x

for $t=1, \dots$

- pick $i = (t \bmod n) + 1$
- sample $x_i^{(t)} \sim p(x_i = \cdot \mid X_{-i}^{(t-1)} = x_{-i}^{(t-1)})$
- set $x_j^{(t)} = x_j^{(t-1)}$ for $j \neq i$

end

in $\{i\}$

time conditional as proposal

⊗ G.S. is M.H. with a time varying proposal

suppose we pick i at time t

for rest to be constant

then proposal is $q_t(x' \mid x^{(t-1)}) = p(x_i' \mid x_{-i}^{(t-1)}) \delta(x_{-i}', x_{-i}^{(t-1)})$

acceptance ratio:

$$\begin{aligned}
 a(x' \mid x^{(t-1)}) &= \frac{p(x_i' \mid x_{-i}') p(x')}{q_t(x' \mid x^{(t-1)}) p(x^{(t-1)})} \rightarrow \frac{p(x_{-i}') p(x_i' \mid x_{-i}')}{p(x_{-i}^{(t-1)}) p(x_i' \mid x_{-i}^{(t-1)})} \\
 &= 1 \quad \parallel \rightarrow \text{always accepts!}
 \end{aligned}$$

convergence of G.S.:

"

Convergence of G.S.:

Let A be Markov transition kernel of one full cycle of G.S (i.e. n steps)

→ homog. M.C.

$$p(x) > 0 \forall x$$

⇒ A is irreducible & aperiodic because $A_{ij} > 0 \forall i, j$
(since can get to any state with n steps)

→ supposing $p(x_i | x_{-i}) > 0$
 $\forall x_i, x_{-i}$

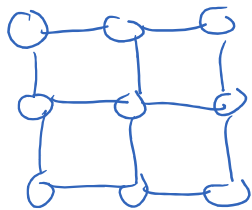
$$\Rightarrow A^t \pi_0 \rightarrow P$$

⊗ also works for random scan (pick $i \in N$ unif. ($1/n$) at each step)

example: G.S. for Ising model

Ising model $x_i \in \{0, 1\}$

UGM:



$$p(x) = \frac{1}{Z(n)} \exp\left(\sum_i m_i x_i + \sum_{\{i,j\} \in E} m_{ij} x_i x_j\right)$$

for G.S.,
want to compute $p(x_i | x_{-i}) \propto p(x_i, x_{-i})$

$$= \exp(m_i x_i + \sum_{j \in N(i)} m_{ij} x_i x_j + \text{rest})$$

⇒ renormalize to get conditional

$$p(x_i=1 | x_{-i}) = \frac{\exp(n_i + \sum_{j \in N(i)} n_{ij} x_j) \cdot \cancel{\exp(\text{rest})}}{1 + \exp(\text{||}) \cdot \cancel{\exp(\text{rest})}}$$
$$= \sigma\left(n_i + \sum_{j \in N(i)} n_{ij} x_j\right)$$